

Least squares estimators for discretely observed stochastic processes driven by small Lévy noises

Hongwei Long^{a,*}, Yasutaka Shimizu^b and Wei Sun^c

^a*Department of Mathematical Sciences, Florida, Atlantic University
Boca Raton, Florida 33431-0991, USA*

^b*Graduate School of Engineering Science, Osaka University
Toyonaka, Osaka 560-8531, Japan*

^c*Department of Mathematics and Statistics, Concordia University
Montreal, Quebec H3G 1M8, Canada*

May 18, 2012

Abstract

We study the problem of parameter estimation for discretely observed stochastic processes driven by additive small Lévy noises. We do not impose any moment condition on the driving Lévy process. Under certain regularity conditions on the drift function, we obtain consistency and rate of convergence of the least squares estimator (LSE) of the drift parameter when a small dispersion coefficient $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$ simultaneously. The asymptotic distribution of the LSE in our general setting is shown to be the convolution of a normal distribution and a distribution related to the jump part of the Lévy process.

Key words: Asymptotic distribution of LSE; consistency of LSE; discrete observations; least squares method; stochastic processes; parameter estimation; small Lévy noises.

MSC2010: Primary 62F12, 62M05; secondary 60G52, 60J75.

1 Introduction

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a basic probability space equipped with a right continuous and increasing family of σ -algebras $(\mathcal{F}_t, t \geq 0)$. Let $(L_t, t \geq 0)$ be a \mathbb{R}^d -valued Lévy process, which is given by

$$L_t = at + \sigma B_t + \int_0^t \int_{|z| \leq 1} z \tilde{N}(ds, dz) + \int_0^t \int_{|z| > 1} z N(ds, dz), \quad (1.1)$$

*Corresponding author. Tel: +1 561 2970810; fax: +1 561 2972436. E-mail address: hlong@fau.edu

where $a = (a_1, \dots, a_d) \in \mathbb{R}^d$, $\sigma = (\sigma_{ij})_{d \times r}$ is a $d \times r$ real-valued matrix, $B_t = (B_t^1, \dots, B_t^r)$ is a r -dimensional standard Brownian motion, $N(ds, dz)$ is an independent Poisson random measure on $\mathbb{R}_+ \times (\mathbb{R}^d \setminus \{0\})$ with characteristic measure $dt\nu(dz)$. Here we assume that $\nu(dz)$ is a Lévy measure on $\mathbb{R}^d \setminus \{0\}$ satisfying $\int_{\mathbb{R}^d \setminus \{0\}} (|z|^2 \wedge 1) \nu(dz) < \infty$ with $|z| = \sqrt{\sum_{i=1}^d z_i^2}$. The stochastic process $X = (X_t, t \geq 0)$, starting from $x_0 \in \mathbb{R}^d$, is defined as the unique strong solution to the following stochastic differential equation (SDE)

$$dX_t = b(X_t, \theta)dt + \varepsilon dL_t, t \in [0, 1]; \quad X_0 = x_0, \quad (1.2)$$

where $\theta \in \Theta = \bar{\Theta}_0$ (the closure of Θ_0) with Θ_0 being an open bounded convex subset of \mathbb{R}^p , and $b = (b_1, \dots, b_d) : \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}^d$ is a known function. Without loss of generality, we assume that $\varepsilon \in (0, 1]$. The regularity conditions on b will be provided in Section 2. Assume that this process is observed at regularly spaced time points $\{t_k = k/n, k = 1, 2, \dots, n\}$. The only unknown quantity in SDE (1.2) is the parameter θ . Let $\theta_0 \in \Theta_0$ be the true value of the parameter θ . The purpose of this paper is to study the least squares estimator for the true value θ_0 based on the sampling data $(X_{t_k})_{k=1}^n$ with small dispersion ε and large sample size n .

In the case of diffusion processes driven by Brownian motion, a popular method is the maximum likelihood estimator (MLE) based on the Girsanov density when the processes can be observed continuously (see Prakasa Rao [29], Liptser and Shiryaev [17], Kutoyants [14]). When a diffusion process is observed only at discrete times, in most cases the transition density and hence the likelihood function of the observations is not explicitly computable. In order to overcome this difficulty, some approximate likelihood methods have been proposed by Lo [18], Pedersen [25]–[26], Poulsen [27], and Aït-Sahalia [1]. For a comprehensive review on MLE and other related methods, we refer to Sørensen [34]. The least squares estimator (LSE) is asymptotically equivalent to the MLE. For the LSE, the convergence in probability was proved in Dorogovcev [3] and Le Breton [16], the strong consistency was studied in Kasonga [10], and the asymptotic distribution was studied in Prakasa Rao [28]. For a more recent comprehensive discussion, we refer to Prakasa Rao [29], Kutoyants [14] and the references therein.

The parametric estimation problems for diffusion processes with jumps based on discrete observations have been studied by Shimizu and Yoshida [32] and Shimizu [30] via the quasi-maximum likelihood. They established consistency and asymptotic normality for the proposed estimators. Moreover, Ogihara and Yoshida [24] showed some stronger results than the ones by Shimizu and Yoshida [32], and also investigated an adaptive Bayes-type estimator with its asymptotic properties. The driving jump processes considered in Shimizu and Yoshida [32], Shimizu [30] and Ogihara and Yoshida [24] include a large class of Lévy processes such as compound Poisson processes, gamma, inverse Gaussian, variance gamma, normal inverse Gaussian or some generalized tempered stable processes. Masuda [22] dealt with the consistency and asymptotic normality of the TFE (trajectory-fitting estimator) and LSE when the driving process is a zero-mean adapted process (including Lévy process) with finite moments. The parametric estimation for Lévy-driven Ornstein-Uhlenbeck processes was also studied by Brockwell *et al.*

[2], Spiliopoulos [36], and Valdivieso *et al.* [43]. However, the aforementioned papers were unable to cover an important class of driving Lévy processes, namely α -stable Lévy motions with $\alpha \in (0, 2)$. Recently, Hu and Long [7]-[8] have started the study on parameter estimation for Ornstein-Uhlenbeck processes driven by α -stable Lévy motions. They obtained some new asymptotic results on the proposed TFE and LSE under continuous or discrete observations, which are different from the classical cases where asymptotic distributions are normal. Fasen [4] extended the results of Hu and Long [8] to multivariate Ornstein-Uhlenbeck processes driven by α -stable Lévy motions. Masuda [23] proposed a self-weighted least absolute deviation estimator for discretely observed ergodic Ornstein-Uhlenbeck processes driven by symmetric Lévy processes.

The asymptotic theory of parametric estimation for diffusion processes with small white noise based on continuous-time observations has been well developed (see, e.g., Kutoyants [12, 13], Yoshida [45, 47], Uchida and Yoshida [41]). There have been many applications of small noise asymptotics to mathematical finance, see for example Yoshida [46], Takahashi [37], Kunitomo and Takahashi [11], Takahashi and Yoshida [38], Uchida and Yoshida [42]. From a practical point of view in parametric inference, it is more realistic and interesting to consider asymptotic estimation for diffusion processes with small noise based on discrete observations. Substantial progress has been made in this direction. Genon-Catalot [5] and Laredo [15] studied the efficient estimation of drift parameters of small diffusions from discrete observations when $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$. Sørensen [33] used martingale estimating functions to establish consistency and asymptotic normality of the estimators of drift and diffusion coefficient parameters when $\varepsilon \rightarrow 0$ and n is fixed. Sørensen and Uchida [35] and Gloter and Sørensen [6] used a contrast function to study the efficient estimation for unknown parameters in both drift and diffusion coefficient functions. Uchida [39, 40] used the martingale estimating function approach to study estimation of drift parameters for small diffusions under weaker conditions. Thus, in the cases of small diffusions, the asymptotic distributions of the estimators are normal under suitable conditions on ε and n .

Long [19] studied the parameter estimation problem for discretely observed one-dimensional Ornstein-Uhlenbeck processes with small Lévy noises. In that paper, the drift function is linear in both x and θ ($b(x, \theta) = -\theta x$), the driving Lévy process is $L_t = aB_t + bZ_t$, where a and b are known constants, $\{B_t, t \geq 0\}$ is the standard Brownian motion and Z_t is a α -stable Lévy motion independent of $\{B_t, t \geq 0\}$. The consistency and rate of convergence of the least squares estimator are established. The asymptotic distribution of the LSE is shown to be the convolution of a normal distribution and a stable distribution. In a similar framework, Long [20] discussed the statistical estimation of the drift parameter for a class of SDEs with special drift function $b(x, \theta) = \theta b(x)$. Ma [21] extended the results of Long [19] to the case when the driving noise is a general Lévy process. However, all the drift functions discussed in Long [19, 20] and Ma [21] are linear in θ , which restricts the applicability of their models and results. In this paper, we allow the drift function $b(x, \theta)$ to be nonlinear in both x and θ , and the driving noise to be a general Lévy process. We are interested in estimating the drift parameter in SDE (1.2) based on discrete observations $\{X_{t_i}\}_{i=1}^n$ when $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$. We shall use the least

squares method to obtain an asymptotically consistent estimator.

Consider the following *contrast function*

$$\Psi_{n,\varepsilon}(\theta) = \sum_{k=1}^n \frac{|X_{t_k} - X_{t_{k-1}} - b(X_{t_{k-1}}, \theta) \cdot \Delta t_{k-1}|^2}{\varepsilon^2 \Delta t_{k-1}},$$

where $\Delta t_{k-1} = t_k - t_{k-1} = 1/n$. Then the LSE $\hat{\theta}_{n,\varepsilon}$ is defined as

$$\hat{\theta}_{n,\varepsilon} := \arg \min_{\theta \in \Theta} \Psi_{n,\varepsilon}(\theta).$$

Since minimizing $\Psi_{n,\varepsilon}(\theta)$ is equivalent to minimizing

$$\Phi_{n,\varepsilon}(\theta) := \varepsilon^2 (\Psi_{n,\varepsilon}(\theta) - \Psi_{n,\varepsilon}(\theta_0)),$$

we may write the LSE as

$$\hat{\theta}_{n,\varepsilon} = \arg \min_{\theta \in \Theta} \Phi_{n,\varepsilon}(\theta).$$

We shall use this fact later for convenience of the proofs.

In the nonlinear case, it is generally very difficult or impossible to obtain an explicit formula for the least squares estimator $\hat{\theta}_{n,\varepsilon}$. However, we can use some nice criteria in statistical inference (see Chapter 5 of Van der Vaart [44] and Shimizu [31] for a more general criterion) to establish the consistency of the LSE as well as its asymptotic behaviors (asymptotic distribution and rate of convergence). In this paper, we consider the asymptotics of the LSE $\hat{\theta}_{n,\varepsilon}$ with high frequency ($n \rightarrow \infty$) and small dispersion ($\varepsilon \rightarrow 0$). Our goal is to prove that $\hat{\theta}_{n,\varepsilon} \rightarrow \theta_0$ in probability and to establish its rate of convergence and asymptotic distributions. We obtain some new asymptotic distributions for the LSE in our general setting, which are the convolutions of normal distribution and a distribution related to the jump part of the driving Lévy process.

The paper is organized as follows. In Section 2, we state our main result with some remarks and examples. We establish the consistency of the LSE $\hat{\theta}_{n,\varepsilon}$, and give its asymptotic distribution, which is a natural extension of the classical small-diffusion cases. All the proofs are given in Section 3.

2 Main results

2.1 Notation and assumptions

Let $X^0 = (X_t^0, t \geq 0)$ be the solution to the underlying ordinary differential equation (ODE) under the true value of the drift parameter:

$$dX_t^0 = b(X_t^0, \theta_0)dt, \quad X_0^0 = x_0.$$

For a multi-index $m = (m_1, \dots, m_k)$, we define a derivative operator in $z \in \mathbb{R}^k$ as $\partial_z^m := \partial_{z_1}^{m_1} \dots \partial_{z_k}^{m_k}$, where $\partial_{z_i}^{m_i} := \partial^{m_i} / \partial z_i^{m_i}$. Let $C^{k,l}(\mathbb{R}^d \times \Theta; \mathbb{R})$ be the space of all functions

$f : \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}$ which is k and l times continuously differentiable with respect to x and θ , respectively. Moreover $C_{\uparrow}^{k,l}(\mathbb{R}^d \times \Theta; \mathbb{R})$ is a class of $f \in C^{k,l}(\mathbb{R}^d \times \Theta; \mathbb{R})$ satisfying that $\sup_{\theta \in \Theta} |\partial_{\theta}^{\alpha} \partial_x^{\beta} f(x, \theta)| \leq C(1 + |x|)^{\lambda}$ for universal positive constants C and λ , where $\alpha = (\alpha_1, \dots, \alpha_p)$ and $\beta = (\beta_1, \dots, \beta_d)$ are multi-indices with $0 \leq \sum_{i=1}^p \alpha_i \leq l$ and $0 \leq \sum_{i=1}^d \beta_i \leq k$, respectively.

We introduce the following set of assumptions.

(A1) There exists a constant $K > 0$ such that

$$|b(x, \theta) - b(y, \theta)| \leq K|x - y|; \quad |b(x, \theta)| \leq K(1 + |x|)$$

for each $x, y \in \mathbb{R}^d$ and $\theta \in \Theta$.

(A2) $b(\cdot, \cdot) \in C_{\uparrow}^{2,3}(\mathbb{R}^d \times \Theta; \mathbb{R})$.

(A3) $\theta \neq \theta_0 \Leftrightarrow b(X_t^0, \theta) \neq b(X_t^0, \theta_0)$ for at least one value of $t \in [0, 1]$.

(A4) $I(\theta_0) = (I^{ij}(\theta_0))_{1 \leq i, j \leq p}$ is positive definite, where

$$I^{ij}(\theta) = \int_0^1 (\partial_{\theta_i} b)^T(X_s^0, \theta) \partial_{\theta_j} b(X_s^0, \theta) ds.$$

It is well-known that SDE (1.2) has a unique strong solution under (A1). For convenience, we shall use C to denote a generic constant whose value may vary from place to place. For a matrix A , we define $|A|^2 = \text{tr}(AA^T)$, where A^T is the transpose of A . In particular, $|\sigma|^2 = \sum_{i=1}^d \sum_{j=1}^r \sigma_{ij}^2$.

2.2 Asymptotic behavior of LSE

The consistency of our estimator $\hat{\theta}_{n,\varepsilon}$ is given as follows.

Theorem 2.1 *Under conditions (A1)–(A3), we have*

$$\hat{\theta}_{n,\varepsilon} \xrightarrow{P_{\theta_0}} \theta_0$$

as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$.

The next theorem gives the asymptotic distribution of $\hat{\theta}_{n,\varepsilon}$. As is easily seen, our result includes the case of Sørensen and Uchida [35] as a special case.

Theorem 2.2 *Under conditions (A1)–(A4), we have*

$$\varepsilon^{-1}(\hat{\theta}_{n,\varepsilon} - \theta_0) \xrightarrow{P_{\theta_0}} I^{-1}(\theta_0)S(\theta_0), \quad (2.1)$$

as $\varepsilon \rightarrow 0$, $n \rightarrow \infty$ and $n\varepsilon \rightarrow \infty$, where

$$S(\theta_0) := \left(\int_0^1 (\partial_{\theta_1} b)^T(X_s^0, \theta_0) dL_s, \dots, \int_0^1 (\partial_{\theta_p} b)^T(X_s^0, \theta_0) dL_s \right)^T.$$

Remark 2.3 One of our main contributions is that we no longer require any high-order moments condition on X as in, e.g., Sørensen and Uchida [35] and others, which makes our results applicable in many practical models.

Remark 2.4 In general, the limiting distribution on the right-hand side of (2.1) is a convolution of a normal distribution and a distribution related to the jump part of the Lévy process. In particular, if the driving Lévy process L is the linear combination of standard Brownian motion and α -stable motion, the limiting distribution becomes the convolution of a normal distribution and a stable distribution.

Remark 2.5 When $d = 1$ and $b(x, \theta) = -\theta x$, i.e., SDE (1.2) is linear and driven by a general Lévy process, Theorem 2.2 reduces to Theorem 1.1 of Ma [21]. When the driving Lévy process is a linear combination of standard Brownian motion and α -stable motion, Theorem 2.2 was discussed in Long [19] and Ma [21].

Remark 2.6 Our results and arguments in the paper can be extended to the SDEs driven by small semi-martingale noises.

Example 2.7 We consider a one-dimensional stochastic process in (1.2) with drift function $b(x, \theta) = \theta_1 + \theta_2 x$. We assume that the true value $\theta_0 = (\theta_1^0, \theta_2^0)$ of $\theta = (\theta_1, \theta_2)$ belongs to $\Theta_0 = (c_1, c_2) \times (c_3, c_4) \subset \mathbb{R}^2$ with $c_1 < c_2$ and $c_3 < c_4$. Then, X^0 satisfies the following ODE

$$dX_t^0 = (\theta_1^0 + \theta_2^0 X_t^0)dt, \quad X_0^0 = x_0.$$

The explicit solution is given by $X_t^0 = e^{\theta_2^0 t} x_0 + \frac{\theta_1^0(e^{\theta_2^0 t} - 1)}{\theta_2^0}$ when $\theta_2^0 \neq 0$; $X_t^0 = x_0 + \theta_1^0 t$ when $\theta_2^0 = 0$. The LSE $\hat{\theta}_{n,\varepsilon} = (\hat{\theta}_{n,\varepsilon,1}, \hat{\theta}_{n,\varepsilon,2})^T$ of θ_0 is given by

$$\begin{aligned} \hat{\theta}_{n,\varepsilon,1} &= (X_1 - X_0) - \hat{\theta}_{n,\varepsilon,2} \left(\frac{1}{n} \sum_{k=1}^n X_{t_{k-1}} \right), \\ \hat{\theta}_{n,\varepsilon,2} &= \frac{\sum_{k=1}^n (X_{t_k} - X_{t_{k-1}}) X_{t_{k-1}} - (X_1 - X_0) \left(\frac{1}{n} \sum_{k=1}^n X_{t_{k-1}} \right)}{\frac{1}{n} \sum_{k=1}^n X_{t_{k-1}}^2 - \left(\frac{1}{n} \sum_{k=1}^n X_{t_{k-1}} \right)^2}. \end{aligned}$$

Note that $\partial_{\theta_1} b(x, \theta) = 1$ and $\partial_{\theta_2} b(x, \theta) = x$. In this case, the limiting random vector in Theorem 2.2 is $I^{-1}(\theta_0) \left(\int_0^1 dL_s, \int_0^1 X_s^0 dL_s \right)^T$, where

$$I(\theta_0) = \begin{pmatrix} \int_0^1 ds & \int_0^1 X_s^0 ds \\ \int_0^1 X_s^0 ds & \int_0^1 (X_s^0)^2 ds \end{pmatrix}.$$

Example 2.8 We consider a one-dimensional stochastic process in (1.2) with drift function $b(x, \theta) = \sqrt{\theta + x^2}$. We assume that the true value θ_0 of θ belongs to $\Theta_0 = (c_1, c_2) \subset \mathbb{R}$ with $0 < c_1 < c_2 < \infty$. Then, X^0 satisfies the following ODE

$$dX_t^0 = \sqrt{\theta_0 + (X_t^0)^2} dt, \quad X_0^0 = x_0.$$

The explicit solution is given by $X_t^0 = \frac{(x_0 + \sqrt{\theta_0 + x_0^2})^2 e^{2t} - \theta_0}{2(x_0 + \sqrt{\theta_0 + x_0^2})e^t}$. It is easy to verify that the LSE $\hat{\theta}_{n,\varepsilon}$ of θ is a solution to the following nonlinear equation

$$\sum_{k=1}^n \frac{X_{t_k} - X_{t_{k-1}}}{\sqrt{\theta + X_{t_{k-1}}^2}} = 1.$$

Since it is impossible to get the explicit expression for $\hat{\theta}_{n,\varepsilon}$, we solve the above equation numerically (e.g. by using Newton's method). Note that $\partial_\theta b(x, \theta) = \frac{1}{2\sqrt{\theta + x^2}}$. It is clear that the limiting random variable in Theorem 2.2 is $I^{-1}(\theta_0) \int_0^1 \frac{1}{2\sqrt{\theta_0 + (X_s^0)^2}} dL_s$, where $I(\theta_0) = \int_0^1 \frac{1}{4(\theta_0 + (X_s^0)^2)} ds$. In particular, we assume that $L_t = aB_t + \sigma Z_t$, where B_t is the standard Brownian motion and Z_t is a standard α -stable Lévy motion independent of B_t . Let us denote by N a random variable with the standard normal distribution and U a random variable with the standard α -stable distribution $S_\alpha(1, \beta, 0)$, where $\alpha \in (0, 2)$ is the index of stability and $\beta \in [-1, 1]$ is the skewness parameter. By using the self-similarity and time change, we can easily show that the limiting random variable in Theorem 2.2 has the identical distribution as

$$aI^{-\frac{1}{2}}(\theta_0)N + \sigma I^{-1}(\theta_0) \left[\int_0^1 \left(\frac{1}{2\sqrt{\theta_0 + (X_s^0)^2}} \right)^\alpha ds \right]^{1/\alpha} U.$$

Example 2.9 We consider a two-dimensional stochastic process in (1.2) with drift function $b(x, \theta) = C + Ax$, where $C = (c_1, c_2)^T$, $A = (A_{ij})_{1 \leq i, j \leq 2}$ and $x = (x_1, x_2)^T$. We assume that the eigenvalues of A have positive real parts. We want to estimate $\theta = (\theta_1, \dots, \theta_6)^T = (c_1, A_{11}, A_{12}, c_2, A_{21}, A_{22})^T \in \Theta \subset \mathbb{R}^6$, whose true value is $\theta_0 = (c_1^0, A_{11}^0, A_{12}^0, c_2^0, A_{21}^0, A_{22}^0)^T$. Then X_t^0 satisfies the following ODE

$$dX_t^0 = (C_0 + A_0 X_t^0) dt, \quad X_0^0 = x_0.$$

The explicit solution is given by $X_t^0 = e^{A_0 t} x_0 + \int_0^t e^{A_0(t-s)} C_0 ds$. After some basic calculation, we find that the LSE $\hat{\theta}_{n,\varepsilon} = (\hat{\theta}_{n,\varepsilon,i})_{1 \leq i \leq 6}$ is given by

$$\begin{pmatrix} \hat{\theta}_{n,\varepsilon,1} \\ \hat{\theta}_{n,\varepsilon,2} \\ \hat{\theta}_{n,\varepsilon,3} \end{pmatrix} = \Lambda_n^{-1} \begin{pmatrix} n \sum_{k=1}^n Y_k^{(1)} \\ n \sum_{k=1}^n Y_k^{(1)} X_{t_{k-1}}^{(1)} \\ n \sum_{k=1}^n Y_k^{(1)} X_{t_{k-1}}^{(2)} \end{pmatrix} \text{ and } \begin{pmatrix} \hat{\theta}_{n,\varepsilon,4} \\ \hat{\theta}_{n,\varepsilon,5} \\ \hat{\theta}_{n,\varepsilon,6} \end{pmatrix} = \Lambda_n^{-1} \begin{pmatrix} n \sum_{k=1}^n Y_k^{(2)} \\ n \sum_{k=1}^n Y_k^{(2)} X_{t_{k-1}}^{(1)} \\ n \sum_{k=1}^n Y_k^{(2)} X_{t_{k-1}}^{(2)} \end{pmatrix},$$

where $X_{t_{k-1}}^{(i)}$ ($i = 1, 2$) are the components of $X_{t_{k-1}}$, $Y_k^{(i)}$ ($i = 1, 2$) are the components of $Y_k = X_{t_k} - X_{t_{k-1}}$, and

$$\Lambda_n = \begin{pmatrix} n & \sum_{k=1}^n X_{t_{k-1}}^{(1)} & \sum_{k=1}^n X_{t_{k-1}}^{(2)} \\ \sum_{k=1}^n X_{t_{k-1}}^{(1)} & \sum_{k=1}^n \left(X_{t_{k-1}}^{(1)} \right)^2 & \sum_{k=1}^n X_{t_{k-1}}^{(1)} X_{t_{k-1}}^{(2)} \\ \sum_{k=1}^n X_{t_{k-1}}^{(2)} & \sum_{k=1}^n X_{t_{k-1}}^{(1)} X_{t_{k-1}}^{(2)} & \sum_{k=1}^n \left(X_{t_{k-1}}^{(2)} \right)^2 \end{pmatrix}.$$

Since it is easy and straightforward to compute the partial derivatives $\partial_{\theta_i} b(x, \theta)$, $1 \leq i \leq 6$, and the limiting random vector in Theorem 2.2, we omit the details here.

3 Proofs

3.1 Proof of Theorem 2.1

We first establish some preliminary lemmas. In the sequel, we shall use the notation

$$Y_t^{n,\varepsilon} := X_{[nt]/n}$$

for the stochastic process X defined by (1.2), where $[nt]$ denotes the integer part of nt .

Lemma 3.1 *The sequence $\{Y_t^{n,\varepsilon}\}$ converges to the deterministic process $\{X_t^0\}$ uniformly on compacts in probability as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$.*

Proof. Note that

$$X_t - X_t^0 = \int_0^t (b(X_s, \theta_0) - b(X_s^0, \theta_0)) ds + \varepsilon L_t. \quad (3.1)$$

By the Lipschitz condition on $b(\cdot)$ in (A1) and the Cauchy-Schwarz inequality, we find that

$$\begin{aligned} |X_t - X_t^0|^2 &\leq 2 \left| \int_0^t (b(X_s, \theta_0) - b(X_s^0, \theta_0)) ds \right|^2 + 2\varepsilon^2 |L_t|^2 \\ &\leq 2t \int_0^t |b(X_s, \theta_0) - b(X_s^0, \theta_0)|^2 ds + 2\varepsilon^2 \sup_{0 \leq s \leq t} |L_s|^2 \\ &\leq 2K^2 t \int_0^t |X_s - X_s^0|^2 ds + 2\varepsilon^2 \sup_{0 \leq s \leq t} |L_s|^2. \end{aligned}$$

By Gronwall's inequality, it follows that

$$|X_t - X_t^0|^2 \leq 2\varepsilon^2 e^{2K^2 t^2} \sup_{0 \leq s \leq t} |L_s|^2$$

and consequently

$$\sup_{0 \leq t \leq T} |X_t - X_t^0| \leq \sqrt{2}\varepsilon e^{K^2 T^2} \sup_{0 \leq t \leq T} |L_t|, \quad (3.2)$$

which goes to zero in probability as $\varepsilon \rightarrow 0$ for each $T > 0$. Since $[nt]/n \rightarrow t$ as $n \rightarrow \infty$, we conclude that the statement holds.

Lemma 3.2 *Let $\tau_m^{n,\varepsilon} = \inf\{t \geq 0 : |X_t^0| \geq m \text{ or } |Y_t^{n,\varepsilon}| \geq m\}$. Then, $\tau_m^{n,\varepsilon} \rightarrow \infty$ a.s. uniformly in n and ε as $m \rightarrow \infty$.*

Proof. Note that

$$X_t = x_0 + \int_0^t b(X_s, \theta_0) ds + \varepsilon L_t.$$

By the linear growth condition on b and the Cauchy-Schwarz inequality, we get

$$\begin{aligned} |X_t|^2 &\leq 2(|x_0| + \varepsilon |L_t|)^2 + 2 \left| \int_0^t b(X_s, \theta_0) ds \right|^2 \\ &\leq 2 \left(|x_0| + \varepsilon \sup_{0 \leq s \leq t} |L_s| \right)^2 + 2t \int_0^t |b(X_s, \theta_0)|^2 ds \\ &\leq 2 \left(|x_0| + \varepsilon \sup_{0 \leq s \leq t} |L_s| \right)^2 + 2K^2 t \int_0^t (1 + |X_s|)^2 ds \\ &\leq \left[2(|x_0| + \varepsilon \sup_{0 \leq s \leq t} |L_s|)^2 + 4K^2 t^2 \right] + 4K^2 t \int_0^t |X_s|^2 ds. \end{aligned}$$

Gronwall's inequality yields that

$$|X_t|^2 \leq \left[2(|x_0| + \varepsilon \sup_{0 \leq s \leq t} |L_s|)^2 + 4K^2 t^2 \right] e^{4K^2 t^2}$$

and

$$|X_t| \leq \left[\sqrt{2}(|x_0| + \varepsilon \sup_{0 \leq s \leq t} |L_s|) + 2Kt \right] e^{2K^2 t^2}.$$

Thus, it follows that

$$|Y_t^{n,\varepsilon}| = |X_{[nt]/n}| \leq \left[\sqrt{2}(|x_0| + \sup_{0 \leq s \leq t} |L_s|) + 2Kt \right] e^{2K^2 t^2},$$

which is almost surely finite. Therefore the proof is complete. \square

We shall use $\nabla_x f(x, \theta) = (\partial_{x_1} f(x, \theta), \dots, \partial_{x_d} f(x, \theta))^T$ to denote the gradient operator of $f(x, \theta)$ with respect to x .

Lemma 3.3 *Let $f \in C_{\uparrow}^{1,1}(\mathbb{R}^d \times \Theta; \mathbb{R})$. Assume (A1)-(A2). Then, we have*

$$\frac{1}{n} \sum_{k=1}^n f(X_{t_{k-1}}, \theta) \xrightarrow{P_{\theta_0}} \int_0^1 f(X_s^0, \theta) ds$$

as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$, uniformly in $\theta \in \Theta$.

Proof. By the differentiability of the function $f(x, \theta)$ and Lemma 3.1, we find that

$$\begin{aligned}
& \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{k=1}^n f(X_{t_{k-1}}, \theta) - \int_0^1 f(X_s^0, \theta) ds \right| \\
&= \sup_{\theta \in \Theta} \left| \int_0^1 f(Y_s^{n,\varepsilon}, \theta) ds - \int_0^1 f(X_s^0, \theta) ds \right| \\
&\leq \sup_{\theta \in \Theta} \int_0^1 |f(Y_s^{n,\varepsilon}, \theta) - f(X_s^0, \theta)| ds \\
&\leq \sup_{\theta \in \Theta} \int_0^1 \left| \int_0^1 (\nabla_x f)^T(X_s^0 + u(Y_s^{n,\varepsilon} - X_s^0), \theta) \cdot (Y_s^{n,\varepsilon} - X_s^0) du \right| ds \\
&\leq \int_0^1 \left(\int_0^1 \sup_{\theta \in \Theta} |\nabla_x f(X_s^0 + u(Y_s^{n,\varepsilon} - X_s^0), \theta)| du \right) |Y_s^{n,\varepsilon} - X_s^0| ds \\
&\leq \int_0^1 C(1 + |X_s^0| + |Y_s^{n,\varepsilon}|)^\lambda |Y_s^{n,\varepsilon} - X_s^0| ds \\
&\leq C \left(1 + \sup_{0 \leq s \leq 1} |X_s^0| + \sup_{0 \leq s \leq 1} |X_s| \right)^\lambda \sup_{0 \leq s \leq 1} |Y_s^{n,\varepsilon} - X_s^0| \\
&\xrightarrow{P_{\theta_0}} 0
\end{aligned}$$

as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$. □

Lemma 3.4 Let $f \in C_{\uparrow}^{1,1}(\mathbb{R}^d \times \Theta; \mathbb{R})$. Assume (A1)-(A2). Then, we have that for each $1 \leq i \leq d$ and each $\theta \in \Theta$,

$$\sum_{k=1}^n f(X_{t_{k-1}}, \theta) (L_{t_k}^i - L_{t_{k-1}}^i) \xrightarrow{P_{\theta_0}} \int_0^1 f(X_s^0, \theta) dL_s^i$$

as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$, where

$$L_t^i = a_i t + \sum_{j=1}^r \sigma_{ij} B_t^j + \int_0^t \int_{|z| \leq 1} z_i \tilde{N}(ds, dz) + \int_0^t \int_{|z| > 1} z_i N(ds, dz)$$

is the i -th component of L_t .

Proof. Note that

$$\sum_{k=1}^n f(X_{t_{k-1}}, \theta) (L_{t_k}^i - L_{t_{k-1}}^i) = \int_0^1 f(Y_s^{n,\varepsilon}, \theta) dL_s^i.$$

Let $\tilde{L}_t^i = L_t^i - \int_0^t \int_{|z| > 1} z_i N(ds, dz)$. Then, we have the following decomposition

$$\int_0^1 f(Y_s^{n,\varepsilon}, \theta) dL_s^i - \int_0^1 f(X_s^0, \theta) dL_s^i = \int_0^1 \int_{|z| > 1} (f(Y_s^{n,\varepsilon}, \theta) - f(X_s^0, \theta)) z_i N(ds, dz)$$

$$+ \int_0^1 (f(Y_s^{n,\varepsilon}, \theta) - f(X_s^0, \theta)) d\tilde{L}_s^i.$$

Similar to the proof of Lemma 3.3, we have

$$\begin{aligned} & \left| \int_0^1 \int_{|z|>1} (f(Y_s^{n,\varepsilon}, \theta) - f(X_s^0, \theta)) z_i N(ds, dz) \right| \\ & \leq \int_0^1 \int_{|z|>1} |f(Y_s^{n,\varepsilon}, \theta) - f(X_s^0, \theta)| |z_i| N(ds, dz) \\ & \leq \int_0^1 \int_{|z|>1} C(1 + |X_s^0| + |Y_s^{n,\varepsilon}|)^\lambda |Y_s^{n,\varepsilon} - X_s^0| |z_i| N(ds, dz) \\ & \leq C \left(1 + \sup_{0 \leq s \leq 1} |X_s^0| + \sup_{0 \leq s \leq 1} |X_s| \right)^\lambda \sup_{0 \leq s \leq 1} |Y_s^{n,\varepsilon} - X_s^0| \int_0^1 \int_{|z|>1} |z_i| N(ds, dz), \end{aligned}$$

which converges to zero in probability as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$ by Lemma 3.1. By using the stopping time $\tau_m^{n,\varepsilon}$, Lemma 3.1, Markov inequality and dominated convergence, we find that for any given $\eta > 0$ and some fixed m

$$\begin{aligned} & P \left(\left| \int_0^1 (f(Y_s^{n,\varepsilon}, \theta) - f(X_s^0, \theta)) 1_{\{s \leq \tau_m^{n,\varepsilon}\}} d\tilde{L}_s^i \right| > \eta \right) \\ & \leq \frac{|a_i|}{\eta} \int_0^1 \mathbb{E} [|f(Y_s^{n,\varepsilon}, \theta) - f(X_s^0, \theta)| 1_{\{s \leq \tau_m^{n,\varepsilon}\}}] ds \\ & \quad + \frac{\sqrt{\sum_{j=1}^r \sigma_{ij}^2}}{\eta} \left(\int_0^1 \mathbb{E} [|f(Y_s^{n,\varepsilon}, \theta) - f(X_s^0, \theta)|^2 1_{\{s \leq \tau_m^{n,\varepsilon}\}}] ds \right)^{1/2} \\ & \quad + \frac{1}{\eta} \left(\int_0^1 \mathbb{E} [|f(Y_s^{n,\varepsilon}, \theta) - f(X_s^0, \theta)|^2 1_{\{s \leq \tau_m^{n,\varepsilon}\}}] ds \cdot \int_{|z| \leq 1} |z_i|^2 \nu(dz) \right)^{1/2}, \quad (3.3) \end{aligned}$$

which goes to zero as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$. Then, we have

$$\begin{aligned} & P \left(\left| \int_0^1 (f(Y_s^{n,\varepsilon}, \theta) - f(X_s^0, \theta)) d\tilde{L}_s^i \right| > \eta \right) \\ & \leq P(\tau_m^{n,\varepsilon} < 1) + P \left(\left| \int_0^1 (f(Y_s^{n,\varepsilon}, \theta) - f(X_s^0, \theta)) 1_{\{s \leq \tau_m^{n,\varepsilon}\}} d\tilde{L}_s^i \right| > \eta \right), \end{aligned}$$

which converges to zero as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$ by Lemma 3.2 and (3.3). This completes the proof. \square

Lemma 3.5 *Let $f \in C_{\uparrow}^{1,1}((\mathbb{R}^d \times \Theta; \mathbb{R}))$. Assume (A1)-(A2). Then, we have that for $1 \leq i \leq d$,*

$$\sum_{k=1}^n f(X_{t_{k-1}}, \theta) (X_{t_k}^i - X_{t_{k-1}}^i - b_i(X_{t_{k-1}}, \theta_0) \Delta t_{k-1}) \xrightarrow{P_{\theta_0}} 0$$

as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$, uniformly in $\theta \in \Theta$, where X_t^i and b_i are the i -th components of X_t and b , respectively.

Proof. Note that

$$X_{t_k}^i = X_{t_{k-1}}^i + \int_{t_{k-1}}^{t_k} b_i(X_s, \theta_0) ds + \varepsilon(L_{t_k}^i - L_{t_{k-1}}^i).$$

It is easy to see that

$$\begin{aligned} & \sum_{k=1}^n f(X_{t_{k-1}}, \theta) (X_{t_k}^i - X_{t_{k-1}}^i - b_i(X_{t_{k-1}}, \theta_0) \Delta t_{k-1}) \\ &= \sum_{k=1}^n \int_{t_{k-1}}^{t_k} f(X_{t_{k-1}}, \theta) (b_i(X_s, \theta_0) - b_i(X_{t_{k-1}}, \theta_0)) ds \\ &+ \varepsilon \sum_{k=1}^n f(X_{t_{k-1}}, \theta) (L_{t_k}^i - L_{t_{k-1}}^i) \\ &= \int_0^1 f(Y_s^{n,\varepsilon}, \theta) (b_i(X_s, \theta_0) - b_i(Y_s^{n,\varepsilon}, \theta_0)) ds + \varepsilon \int_0^1 f(Y_s^{n,\varepsilon}, \theta) dL_s^i. \end{aligned}$$

By the given condition on f and the Lipschitz condition on b , we have

$$\begin{aligned} & \sup_{\theta \in \Theta} \left| \int_0^1 f(Y_s^{n,\varepsilon}, \theta) (b_i(X_s, \theta_0) - b_i(Y_s^{n,\varepsilon}, \theta_0)) ds \right| \\ & \leq \int_0^1 \sup_{\theta \in \Theta} |f(Y_s^{n,\varepsilon}, \theta)| \cdot K |X_s - Y_s^{n,\varepsilon}| ds \\ & \leq KC \int_0^1 (1 + |Y_s^{n,\varepsilon}|)^\lambda (|X_s - X_s^0| + |Y_s^{n,\varepsilon} - X_s^0|) ds \\ & \leq KC \left(1 + \sup_{0 \leq t \leq 1} |X_t| \right)^\lambda \left(\sup_{0 \leq s \leq 1} |X_s - X_s^0| + \sup_{0 \leq s \leq 1} |Y_s^{n,\varepsilon} - X_s^0| \right), \end{aligned}$$

which converges to zero in probability as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$ by Lemma 3.1. Next using the decomposition of L_t , we have

$$\begin{aligned} & \sup_{\theta \in \Theta} \left| \varepsilon \int_0^1 f(Y_s^{n,\varepsilon}, \theta) dL_s^i \right| \\ & \leq \varepsilon \sup_{\theta \in \Theta} \left| a_i \int_0^1 f(Y_s^{n,\varepsilon}, \theta) ds \right| + \varepsilon \sup_{\theta \in \Theta} \left| \int_0^1 f(Y_s^{n,\varepsilon}, \theta) \sum_{j=1}^r \sigma_{ij} dB_s^j \right| \\ & + \varepsilon \sup_{\theta \in \Theta} \left| \int_0^1 \int_{|z| \leq 1} f(Y_s^{n,\varepsilon}, \theta) z_i \tilde{N}(ds, dz) \right| \\ & + \varepsilon \sup_{\theta \in \Theta} \left| \int_0^1 \int_{|z| > 1} f(Y_s^{n,\varepsilon}, \theta) z_i N(ds, dz) \right|. \end{aligned}$$

It is clear that

$$\begin{aligned} \varepsilon \sup_{\theta \in \Theta} \left| a_i \int_0^1 f(Y_s^{n,\varepsilon}, \theta) ds \right| &\leq \varepsilon |a_i| C \int_0^1 (1 + |Y_s^{n,\varepsilon}|)^\lambda ds \\ &\leq \varepsilon |a_i| C \left(1 + \sup_{0 \leq s \leq 1} |X_s| \right)^\lambda, \end{aligned}$$

which converges to zero in probability as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$, and

$$\begin{aligned} \varepsilon \sup_{\theta \in \Theta} \left| \int_0^1 \int_{|z|>1} f(Y_s^{n,\varepsilon}, \theta) z_i N(ds, dz) \right| &\leq \varepsilon \int_0^1 \int_{|z|>1} \sup_{\theta \in \Theta} |f(Y_s^{n,\varepsilon}, \theta)| \cdot |z_i| N(ds, dz) \\ &\leq \varepsilon \int_0^1 \int_{|z|>1} C(1 + |Y_s^{n,\varepsilon}|)^\lambda \cdot |z_i| N(ds, dz) \\ &\leq \varepsilon C \left(1 + \sup_{0 \leq s \leq 1} |X_s| \right)^\lambda \int_0^1 \int_{|z|>1} |z_i| N(ds, dz), \end{aligned}$$

which converges to zero in probability. Note that

$$\begin{aligned} &P \left(\varepsilon \sup_{\theta \in \Theta} \left| \int_0^1 f(Y_s^{n,\varepsilon}, \theta) \sum_{j=1}^r \sigma_{ij} dB_s^j \right| > \eta \right) \\ &\leq P(\tau_m^{n,\varepsilon} < 1) + P \left(\varepsilon \sup_{\theta \in \Theta} \left| \int_0^1 f(Y_s^{n,\varepsilon}, \theta) 1_{\{s \leq \tau_m^{n,\varepsilon}\}} \sum_{j=1}^r \sigma_{ij} dB_s^j \right| > \eta \right). \end{aligned} \quad (3.4)$$

Let

$$u_{n,\varepsilon}^i(\theta) = \varepsilon \int_0^1 f(Y_s^{n,\varepsilon}, \theta) 1_{\{s \leq \tau_m^{n,\varepsilon}\}} \sum_{j=1}^r \sigma_{ij} dB_s^j, \quad 1 \leq i \leq d.$$

We want to prove that $u_{n,\varepsilon}^i(\theta) \rightarrow 0$ in probability as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$, uniformly in $\theta \in \Theta$. It suffices to show the pointwise convergence and the tightness of the sequence $\{u_{n,\varepsilon}^i(\cdot)\}$. For the pointwise convergence, by the Chebyshev inequality and Ito's isometry, we have

$$\begin{aligned} &P(|u_{n,\varepsilon}^i(\theta)| > \eta) \\ &\leq \varepsilon^2 \eta^{-2} \mathbb{E} \left[\left| \int_0^1 f(Y_s^{n,\varepsilon}, \theta) 1_{\{s \leq \tau_m^{n,\varepsilon}\}} \sum_{j=1}^r \sigma_{ij} dB_s^j \right|^2 \right] \\ &\leq \left(\sum_{j=1}^r \sigma_{ij}^2 \right) \varepsilon^2 \eta^{-2} \int_0^1 \mathbb{E} [|f(Y_s^{n,\varepsilon}, \theta)|^2 1_{\{s \leq \tau_m^{n,\varepsilon}\}}] ds \end{aligned}$$

$$\begin{aligned}
&\leq \left(\sum_{j=1}^r \sigma_{ij}^2 \right) \varepsilon^2 \eta^{-2} \int_0^1 \mathbb{E} \left[C^2 (1 + |Y_s^{n,\varepsilon}|)^{2\lambda} 1_{\{s \leq \tau_m^{n,\varepsilon}\}} \right] ds \\
&\leq \left(\sum_{j=1}^r \sigma_{ij}^2 \right) \varepsilon^2 \eta^{-2} C^2 (1 + m)^{2\lambda},
\end{aligned} \tag{3.5}$$

which converges to zero as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$ with fixed m . For the tightness of $\{u_{n,\varepsilon}^i(\cdot)\}$, by using Theorem 20 in Appendix I of Ibragimov and Has'minskii [9], it is enough to prove the following two inequalities

$$\mathbb{E}[|u_{n,\varepsilon}^i(\theta)|^{2q}] \leq C, \tag{3.6}$$

$$\mathbb{E}[|u_{n,\varepsilon}^i(\theta_2) - u_{n,\varepsilon}^i(\theta_1)|^{2q}] \leq C|\theta_2 - \theta_1|^{2q} \tag{3.7}$$

for $\theta, \theta_1, \theta_2 \in \Theta$, where $2q > p$. The proof of (3.6) is very similar to moment estimates in (3.5) by replacing Ito's isometry with the Burkholder-Davis-Gundy inequality. So we omit the details here. For (3.7), by using Taylor's formula and the Burkholder-Davis-Gundy inequality, we have

$$\begin{aligned}
&\mathbb{E}[|u_{n,\varepsilon}^i(\theta_2) - u_{n,\varepsilon}^i(\theta_1)|^{2q}] \\
&\leq \varepsilon^{2q} C_q \left(\sum_{j=1}^r \sigma_{ij}^2 \right)^q \mathbb{E} \left[\left(\int_0^1 (f(Y_s^{n,\varepsilon}, \theta_2) - f(Y_s^{n,\varepsilon}, \theta_1))^2 1_{\{s \leq \tau_m^{n,\varepsilon}\}} ds \right)^q \right] \\
&\leq \varepsilon^{2q} C_q \left(\sum_{j=1}^r \sigma_{ij}^2 \right)^q \mathbb{E} \left[\left(\int_0^1 \int_0^1 |\theta_2 - \theta_1|^2 |\nabla_\theta f(Y_s^{n,\varepsilon}, \theta_1 + v(\theta_2 - \theta_1))|^2 1_{\{s \leq \tau_m^{n,\varepsilon}\}} dv ds \right)^q \right] \\
&\leq \varepsilon^{2q} C_q \left(\sum_{j=1}^r \sigma_{ij}^2 \right)^q C^{2q} |\theta_2 - \theta_1|^{2q} \mathbb{E} \left[\left(\int_0^1 (1 + |Y_s^{n,\varepsilon}|)^{2\lambda} 1_{\{s \leq \tau_m^{n,\varepsilon}\}} ds \right)^q \right] \\
&\leq \varepsilon^{2q} C_q \left(\sum_{j=1}^r \sigma_{ij}^2 \right)^q C^{2q} (1 + m)^{2\lambda q} |\theta_2 - \theta_1|^{2q}.
\end{aligned}$$

Combining (3.4) and the above arguments, we have that $\varepsilon \sup_{\theta \in \Theta} \left| \int_0^1 f(Y_s^{n,\varepsilon}, \theta) \sum_{j=1}^r \sigma_{ij} dB_s^j \right|$ converges to zero in probability as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$. Similarly, we can prove that $\varepsilon \sup_{\theta \in \Theta} \left| \int_0^1 \int_{|z| \leq 1} f(Y_s^{n,\varepsilon}, \theta) z_i \tilde{N}(ds, dz) \right|$ converges to zero in probability as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$. Therefore, the proof is complete. \square

Now we are in a position to prove Theorem 2.1.

Proof of Theorem 2.1. Note that

$$\Phi_{n,\varepsilon}(\theta) = -2 \sum_{k=1}^n (b(X_{t_{k-1}}, \theta) - b(X_{t_{k-1}}, \theta_0))^T (X_{t_k} - X_{t_{k-1}} - n^{-1} b(X_{t_{k-1}}, \theta_0))$$

$$\begin{aligned}
& + \frac{1}{n} \sum_{k=1}^n |b(X_{t_{k-1}}, \theta) - b(X_{t_{k-1}}, \theta_0)|^2. \\
& := \Phi_{n,\varepsilon}^{(1)}(\theta) + \Phi_{n,\varepsilon}^{(2)}(\theta).
\end{aligned}$$

By Lemma 3.5 and let $f(x, \theta) = b_i(x, \theta) - b_i(x, \theta_0)$ ($1 \leq i \leq d$), we have $\sup_{\theta \in \Theta} |\Phi_{n,\varepsilon}^{(1)}(\theta)| \xrightarrow{P_{\theta_0}} 0$ as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$. By using Lemma 3.3 with $f(x, \theta) = |b(x, \theta) - b(x, \theta_0)|^2$, we find $\sup_{\theta \in \Theta} |\Phi_{n,\varepsilon}^{(2)}(\theta) - F(\theta)| \xrightarrow{P_{\theta_0}} 0$ as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$, where $F(\theta) = \int_0^1 |b(X_t^0, \theta) - b(X_t^0, \theta_0)|^2 dt$. Thus combining the previous arguments, we have

$$\sup_{\theta \in \Theta} |\Phi_{n,\varepsilon}(\theta) - F(\theta)| \xrightarrow{P_{\theta_0}} 0$$

as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$, and that (A3) and the continuity of X^0 yield that

$$\inf_{|\theta - \theta_0| > \delta} F(\theta) > F(\theta_0) = 0,$$

for each $\delta > 0$. Therefore, by Theorem 5.9 of van der Vaart [44], we have the desired consistency, i.e., $\hat{\theta}_{n,\varepsilon} \xrightarrow{P_{\theta_0}} \theta_0$ as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$. This completes the proof. \square

3.2 Proof of Theorem 2.2

Note that

$$\nabla_{\theta} \Phi_{n,\varepsilon}(\theta) = -2 \sum_{k=1}^n (\nabla_{\theta} b)^T(X_{t_{k-1}}, \theta) (X_{t_k} - X_{t_{k-1}} - b(X_{t_{k-1}}, \theta) \Delta t_{k-1}).$$

Let $G_{n,\varepsilon}(\theta) = (G_{n,\varepsilon}^1, \dots, G_{n,\varepsilon}^p)^T$ with

$$G_{n,\varepsilon}^i(\theta) = \sum_{k=1}^n (\partial_{\theta_i} b)^T(X_{t_{k-1}}, \theta) (X_{t_k} - X_{t_{k-1}} - b(X_{t_{k-1}}, \theta) \Delta t_{k-1}), \quad i = 1, \dots, p,$$

and let $K_{n,\varepsilon}(\theta) = \nabla_{\theta} G_{n,\varepsilon}(\theta)$, which is a $p \times p$ matrix consisting of elements $K_{n,\varepsilon}^{ij}(\theta) = \partial_{\theta_j} G_{n,\varepsilon}^i(\theta)$, $1 \leq i, j \leq p$. Moreover, we introduce the following function

$$K^{ij}(\theta) = \int_0^1 (\partial_{\theta_j} \partial_{\theta_i} b)^T(X_s^0, \theta) (b(X_s^0, \theta_0) - b(X_s^0, \theta)) ds - I^{ij}(\theta), \quad 1 \leq i, j \leq p.$$

Then we define the matrix function $K(\theta) = (K^{ij}(\theta))_{1 \leq i, j \leq p}$.

Before proving Theorem 2.2, we prepare some preliminary results.

Lemma 3.6 *Assume (A1)-(A2). Then, we have that for each $i = 1, \dots, p$*

$$\varepsilon^{-1} G_{n,\varepsilon}^i(\theta_0) \xrightarrow{P_{\theta_0}} \int_0^1 (\partial_{\theta_i} b)^T(X_s^0, \theta_0) dL_s$$

as $\varepsilon \rightarrow 0$, $n \rightarrow \infty$ and $n\varepsilon \rightarrow \infty$.

Proof. Note that for $1 \leq i \leq p$

$$\begin{aligned}
\varepsilon^{-1} G_{n,\varepsilon}^i(\theta_0) &= \varepsilon^{-1} \sum_{k=1}^n (\partial_{\theta_i} b)^T(X_{t_{k-1}}, \theta_0) (X_{t_k} - X_{t_{k-1}} - b(X_{t_{k-1}}, \theta_0) \Delta t_{k-1}) \\
&= \varepsilon^{-1} \sum_{k=1}^n (\partial_{\theta_i} b)^T(X_{t_{k-1}}, \theta_0) \int_{t_{k-1}}^{t_k} (b(X_s, \theta_0) - b(X_{t_{k-1}}, \theta_0)) ds \\
&\quad + \sum_{k=1}^n (\partial_{\theta_i} b)^T(X_{t_{k-1}}, \theta_0) (L_{t_k} - L_{t_{k-1}}) \\
&:= H_{n,\varepsilon}^{(1)}(\theta_0) + H_{n,\varepsilon}^{(2)}(\theta_0).
\end{aligned}$$

By using Lemma 3.4 and letting $f(x, \theta) = \partial_{\theta_i} b_j(x, \theta)$ ($1 \leq i \leq p$, $1 \leq j \leq d$) with $\theta = \theta_0$, we have

$$H_{n,\varepsilon}^{(2)}(\theta_0) = \int_0^1 (\partial_{\theta_i} b)^T(Y_s^{n,\varepsilon}, \theta_0) dL_s \xrightarrow{P_{\theta_0}} \int_0^1 (\partial_{\theta_i} b)^T(X_s^0, \theta_0) dL_s$$

as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$. It suffices to prove that $H_{n,\varepsilon}^{(1)}(\theta_0)$ converges to zero in probability. For $H_{n,\varepsilon}^{(1)}(\theta_0)$, we need some delicate estimate for the process X_t . For $s \in [t_{k-1}, t_k]$, we have

$$X_s - X_{t_{k-1}} = \int_{t_{k-1}}^s (b(X_u, \theta_0) - b(X_{t_{k-1}}, \theta_0)) du + b(X_{t_{k-1}}, \theta_0)(s - t_{k-1}) + \varepsilon(L_s - L_{t_{k-1}}).$$

By the Lipschitz condition on b and the Cauchy-Schwarz inequality, we find that

$$\begin{aligned}
|X_s - X_{t_{k-1}}|^2 &\leq 2 \left| \int_{t_{k-1}}^s (b(X_u, \theta_0) - b(X_{t_{k-1}}, \theta_0)) du \right|^2 \\
&\quad + 2 (|b(X_{t_{k-1}}, \theta_0)| (s - t_{k-1}) + \varepsilon |L_s - L_{t_{k-1}}|)^2 \\
&\leq 2K^2 n^{-1} \int_{t_{k-1}}^s |X_u - X_{t_{k-1}}|^2 du \\
&\quad + 2 \left(n^{-1} |b(X_{t_{k-1}}, \theta_0)| + \varepsilon \sup_{t_{k-1} \leq s \leq t_k} |L_s - L_{t_{k-1}}| \right)^2.
\end{aligned}$$

By Gronwall's inequality, we get

$$|X_s - X_{t_{k-1}}|^2 \leq 2 \left(n^{-1} |b(X_{t_{k-1}}, \theta_0)| + \varepsilon \sup_{t_{k-1} \leq s \leq t_k} |L_s - L_{t_{k-1}}| \right)^2 e^{2K^2 n^{-1} (s - t_{k-1})}.$$

It further follows that

$$\sup_{t_{k-1} \leq s \leq t_k} |X_s - X_{t_{k-1}}| \leq \sqrt{2} \left(n^{-1} |b(X_{t_{k-1}}, \theta_0)| + \varepsilon \sup_{t_{k-1} \leq s \leq t_k} |L_s - L_{t_{k-1}}| \right) e^{K^2/n^2} \quad (3.8)$$

Thus, by the Lipschitz condition on b and (3.8), we get

$$\begin{aligned}
|H_{n,\varepsilon}^{(1)}(\theta_0)| &\leq \varepsilon^{-1} \sum_{k=1}^n |\partial_{\theta_i} b(X_{t_{k-1}}, \theta_0)| \cdot \left| \int_{t_{k-1}}^{t_k} (b(X_s, \theta_0) - b(X_{t_{k-1}}, \theta_0)) ds \right| \\
&\leq \varepsilon^{-1} \sum_{k=1}^n |\partial_{\theta_i} b(X_{t_{k-1}}, \theta_0)| \cdot \int_{t_{k-1}}^{t_k} |b(X_s, \theta_0) - b(X_{t_{k-1}}, \theta_0)| ds \\
&\leq \varepsilon^{-1} \sum_{k=1}^n |\partial_{\theta_i} b(X_{t_{k-1}}, \theta_0)| \int_{t_{k-1}}^{t_k} K |X_s - X_{t_{k-1}}| ds \\
&\leq (n\varepsilon)^{-1} K \sum_{k=1}^n |\partial_{\theta_i} b(X_{t_{k-1}}, \theta_0)| \sup_{t_{k-1} \leq s \leq t_k} |X_s - X_{t_{k-1}}| \\
&\leq \frac{\sqrt{2} K e^{K^2/n^2}}{n\varepsilon} \cdot \frac{1}{n} \sum_{k=1}^n |\partial_{\theta_i} b(X_{t_{k-1}}, \theta_0)| \cdot |b(X_{t_{k-1}}, \theta_0)| \\
&\quad + \frac{\sqrt{2} K e^{K^2/n^2}}{n} \sum_{k=1}^n |\partial_{\theta_i} b(X_{t_{k-1}}, \theta_0)| \sup_{t_{k-1} \leq s \leq t_k} |L_s - L_{t_{k-1}}| \\
&:= H_{n,\varepsilon}^{(1,1)}(\theta_0) + H_{n,\varepsilon}^{(1,2)}(\theta_0).
\end{aligned}$$

It is easy to see that $H_{n,\varepsilon}^{(1,1)}(\theta_0)$ converges to zero in probability as $n\varepsilon \rightarrow \infty$ since

$$\frac{1}{n} \sum_{k=1}^n |\partial_{\theta_i} b(X_{t_{k-1}}, \theta_0)| \cdot |b(X_{t_{k-1}}, \theta_0)| \leq CK \left(1 + \sup_{0 \leq s \leq 1} |X_s| \right)^{\lambda+1} < \infty \quad a.s.$$

(cf. (3.2)). By using the basic fact that

$$\frac{1}{n} \sum_{k=1}^n \sup_{t_{k-1} \leq s \leq t_k} |L_s - L_{t_{k-1}}| = o_P(1),$$

we find that

$$H_{n,\varepsilon}^{(1,2)}(\theta_0) \leq \sqrt{2} K e^{K^2/n^2} C \left(1 + \sup_{0 \leq s \leq 1} |X_s| \right)^{\lambda} \frac{1}{n} \sum_{k=1}^n \sup_{t_{k-1} \leq s \leq t_k} |L_s - L_{t_{k-1}}|,$$

which converges to zero in probability as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$. Therefore the proof is complete. \square

Lemma 3.7 *Assume (A1)-(A4). Then, we have*

$$\sup_{\theta \in \Theta} |K_{n,\varepsilon}(\theta) - K(\theta)| \xrightarrow{P_{\theta_0}} 0$$

as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$.

Proof. It suffices to prove that for $1 \leq i, j \leq p$

$$\sup_{\theta \in \Theta} |K_{n,\varepsilon}^{ij}(\theta) - K^{ij}(\theta)| \xrightarrow{P_{\theta_0}} 0$$

as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$. Note that

$$\begin{aligned} K_{n,\varepsilon}^{ij}(\theta) &= \partial_{\theta_j} G_{n,\varepsilon}^i(\theta) \\ &= \sum_{k=1}^n (\partial_{\theta_j} \partial_{\theta_i} b)^T(X_{t_{k-1}}, \theta) (X_{t_k} - X_{t_{k-1}} - b(X_{t_{k-1}}, \theta_0) \Delta t_{k-1}) \\ &\quad + \frac{1}{n} \sum_{i=1}^n [(\partial_{\theta_j} \partial_{\theta_i} b)^T(X_{t_{k-1}}, \theta) (b(X_{t_{k-1}}, \theta_0) - b(X_{t_{k-1}}, \theta)) \\ &\quad - (\partial_{\theta_i} b)^T(X_{t_{k-1}}, \theta) \partial_{\theta_j} b(X_{t_{k-1}}, \theta)] \\ &:= K_{n,\varepsilon}^{ij,(1)}(\theta) + K_{n,\varepsilon}^{ij,(2)}(\theta). \end{aligned}$$

By using Lemma 3.5 and letting $f(x, \theta) = \partial_{\theta_j} \partial_{\theta_i} b_l(x, \theta)$ ($1 \leq i, j \leq p, 1 \leq l \leq d$), we have that $\sup_{\theta \in \Theta} |K_{n,\varepsilon}^{ij,(1)}(\theta)|$ converges to zero in probability as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$. By using Lemma 3.3 and letting $f(x, \theta) = (\partial_{\theta_j} \partial_{\theta_i} b)^T(x, \theta) (b(x, \theta_0) - b(x, \theta)) - (\partial_{\theta_i} b)^T(x, \theta) \partial_{\theta_j} b(x, \theta)$, it follows that $\sup_{\theta \in \Theta} |K_{n,\varepsilon}^{ij,(2)}(\theta) - K^{ij}(\theta)|$ converges to zero in probability as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$. Thus, the proof is complete. \square

Finally we are ready to prove Theorem 2.2.

Proof of Theorem 2.2. The proof ideas mainly follow Uchida [39]. Let $B(\theta_0; \rho) = \{\theta : |\theta - \theta_0| \leq \rho\}$ for $\rho > 0$. Then, by the consistency of $\hat{\theta}_{n,\varepsilon}$, there exists a sequence $\eta_{n,\varepsilon} \rightarrow 0$ as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$ such that $B(\theta_0; \eta_{n,\varepsilon}) \subset \Theta_0$, and that $P_{\theta_0}[\hat{\theta}_{n,\varepsilon} \in B(\theta_0; \eta_{n,\varepsilon})] \rightarrow 1$. When $\hat{\theta}_{n,\varepsilon} \in B(\theta_0; \eta_{n,\varepsilon})$, it follows by Taylor's formula that

$$D_{n,\varepsilon} S_{n,\varepsilon} = \varepsilon^{-1} G_{n,\varepsilon}(\hat{\theta}_{n,\varepsilon}) - \varepsilon^{-1} G_{n,\varepsilon}(\theta_0),$$

where $D_{n,\varepsilon} = \int_0^1 K_{n,\varepsilon}(\theta_0 + u(\hat{\theta}_{n,\varepsilon} - \theta_0)) du$ and $S_{n,\varepsilon} = \varepsilon^{-1}(\hat{\theta}_{n,\varepsilon} - \theta_0)$ since $B(\theta_0; \eta_{n,\varepsilon})$ is a convex subset of Θ_0 . We have

$$\begin{aligned} |D_{n,\varepsilon} - K_{n,\varepsilon}(\theta_0)| 1_{\{\hat{\theta}_{n,\varepsilon} \in B(\theta_0; \eta_{n,\varepsilon})\}} &\leq \sup_{\theta \in B(\theta_0; \eta_{n,\varepsilon})} |K_{n,\varepsilon}(\theta) - K_{n,\varepsilon}(\theta_0)| \\ &\leq \sup_{\theta \in B(\theta_0; \eta_{n,\varepsilon})} |K_{n,\varepsilon}(\theta) - K(\theta)| \\ &\quad + \sup_{\theta \in B(\theta_0; \eta_{n,\varepsilon})} |K(\theta) - K(\theta_0)| + |K_{n,\varepsilon}(\theta_0) - K(\theta_0)|. \end{aligned}$$

Consequently, it follows from Lemma 3.7 that

$$D_{n,\varepsilon} \xrightarrow{P_{\theta_0}} K(\theta_0), \quad \varepsilon \rightarrow 0, \quad n \rightarrow \infty.$$

Note that $K(\theta)$ is continuous with respect to θ . Since $-K(\theta_0) = I(\theta_0)$ is positive definite, there exists a positive constant $\delta > 0$ such that $\inf_{|w|=1} |K(\theta_0)w| > 2\delta$. For such a $\delta > 0$, there exists $\varepsilon(\delta) > 0$ and $N(\delta) \in \mathbb{N}$ such that for any $\varepsilon \in (0, \varepsilon(\delta))$, $n > N(\delta)$, $B(\theta_0; \eta_{n,\varepsilon}) \subset \Theta_0$ and $|K(\theta) - K(\theta_0)| < \delta/2$ for $\theta \in B(\theta_0; \eta_{n,\varepsilon})$. For such $\delta > 0$, let

$$\Gamma_{n,\varepsilon} = \left\{ \sup_{|\theta - \theta_0| < \eta_{n,\varepsilon}} |K_{n,\varepsilon}(\theta) - K(\theta_0)| < \frac{\delta}{2}, \hat{\theta}_{n,\varepsilon} \in B(\theta_0; \eta_{n,\varepsilon}) \right\}.$$

Then, for any $\varepsilon \in (0, \varepsilon(\delta))$ and $n > N(\delta)$, we have, on $\Gamma_{n,\varepsilon}$,

$$\begin{aligned} \sup_{|w|=1} |(D_{n,\varepsilon} - K(\theta_0))w| &\leq \sup_{|w|=1} \left| \left(D_{n,\varepsilon} - \int_0^1 K(\theta_0 + u(\hat{\theta}_{n,\varepsilon} - \theta_0)) du \right) w \right| \\ &\quad + \sup_{|w|=1} \left| \left(\int_0^1 K(\theta_0 + u(\hat{\theta}_{n,\varepsilon} - \theta_0)) du - K(\theta_0) \right) w \right| \\ &\leq \sup_{|\theta - \theta_0| \leq \eta_{n,\varepsilon}} |K_{n,\varepsilon}(\theta) - K(\theta)| + \frac{\delta}{2} < \delta. \end{aligned}$$

Thus, on $\Gamma_{n,\varepsilon}$,

$$\inf_{|w|=1} |D_{n,\varepsilon}w| \geq \inf_{|w|=1} |K(\theta_0)w| - \sup_{|w|=1} |(D_{n,\varepsilon} - K(\theta_0))w| > 2\delta - \delta = \delta > 0.$$

Hence, letting

$$\mathcal{D}_{n,\varepsilon} = \{D_{n,\varepsilon} \text{ is invertible}, \hat{\theta}_{n,\varepsilon} \in B(\theta_0; \eta_{n,\varepsilon})\},$$

we see that $P_{\theta_0}[\mathcal{D}_{n,\varepsilon}] \geq P_{\theta_0}[\Gamma_{n,\varepsilon}] \rightarrow 1$ as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$ by Lemma 3.7. Now set

$$U_{n,\varepsilon} = D_{n,\varepsilon} 1_{\mathcal{D}_{n,\varepsilon}} + I_{p \times p} 1_{\mathcal{D}_{n,\varepsilon}^c},$$

where $I_{p \times p}$ is the identity matrix. Then it is easy to see that

$$|U_{n,\varepsilon} - K(\theta_0)| \leq |D_{n,\varepsilon} - K(\theta_0)| 1_{\mathcal{D}_{n,\varepsilon}} + |I_{p \times p} - K(\theta_0)| 1_{\mathcal{D}_{n,\varepsilon}^c} \xrightarrow{P_{\theta_0}} 0,$$

since $P_{\theta_0}[\mathcal{D}_{n,\varepsilon}] \rightarrow 1$. Thus, by Lemma 3.6, we obtain that

$$\begin{aligned} S_{n,\varepsilon} &= U_{n,\varepsilon}^{-1} D_{n,\varepsilon} S_{n,\varepsilon} 1_{\mathcal{D}_{n,\varepsilon}} + S_{n,\varepsilon} 1_{\mathcal{D}_{n,\varepsilon}^c} \\ &= U_{n,\varepsilon}^{-1} (-\varepsilon^{-1} G_{n,\varepsilon}(\theta_0)) 1_{\mathcal{D}_{n,\varepsilon}} + S_{n,\varepsilon} 1_{\mathcal{D}_{n,\varepsilon}^c} \\ &\xrightarrow{P_{\theta_0}} (I(\theta_0))^{-1} \left(\int_0^1 (\partial_{\theta_1} b)^T(X_s^0, \theta_0) dL_s, \dots, \int_0^1 (\partial_{\theta_p} b)^T(X_s^0, \theta_0) dL_s \right)^T \end{aligned}$$

as $\varepsilon \rightarrow 0$, $n \rightarrow \infty$ and $n\varepsilon \rightarrow \infty$. This completes the proof. \square

References

- [1] Y. Aït-Sahalia, Maximum likelihood estimation of discretely sampled diffusion: a closed-form approximation approach, *Econometrica* **70** (2002) 223-262.
- [2] P.J. Brockwell, R.A. Davis, Y. Yang, Estimation for non-negative Lévy-driven Ornstein-Uhlenbeck processes, *J. Appl. Probab.* **44** (2007) 977-989.
- [3] A. Ja. Dorogovcev, The consistency of an estimate of a parameter of a stochastic differential equation, *Theory Probab. Math. Stat.* **10** (1976) 73-82.
- [4] V. Fasen, Statistical estimation of multivariate Ornstein-Uhlenbeck processes and applications to co-integration, To appear in *J. Econometrics*, 2011.
- [5] V. Genon-Catalot, Maximum contrast estimation for diffusion processes from discrete observations, *Statistics* **21** (1990) 99-116.
- [6] A. Gloter, M. Sørensen, Estimation for stochastic differential equations with a small diffusion coefficient, *Stochastic Process. Appl.* **119** (2009) 679-699.
- [7] Y. Hu, H. Long, Parameter estimation for Ornstein-Uhlenbeck processes driven by α -stable Lévy motions, *Communications on Stochastic Analysis* **1** (2007) 175-192.
- [8] Y. Hu, H. Long, Least squares estimator for Ornstein-Uhlenbeck processes driven by α -stable motions, *Stochastic Process. Appl.* **119** (2009) 2465-2480.
- [9] I.A. Ibragimov, R.Z. Has'minskii, *Statistical Estimation: Asymptotic Theory*, Springer-Verlag, New York, Berlin, 1981.
- [10] R.A. Kasonga, The consistency of a nonlinear least squares estimator for diffusion processes, *Stochastic Process. Appl.* **30** (1988) 263-275.
- [11] N. Kunitomo, A. Takahashi, The asymptotic expansion approach to the valuation of interest rate contingent claims, *Math. Finance* **11** (2001) 117-151.
- [12] Yu.A. Kutoyants, *Parameter Estimation for Stochastic Processes*, Heldermann, Berlin, 1984.
- [13] Yu.A. Kutoyants, *Identification of Dynamical Systems with Small Noise*, Kluwer, Dordrecht, 1994.
- [14] Yu.A. Kutoyants, *Statistical Inference for Ergodic Diffusion Processes*, Springer-Verlag, London, Berlin, Heidelberg, 2004.
- [15] C.F. Laredo, A sufficient condition for asymptotic sufficiency of incomplete observations of a diffusion process, *Ann. Statist.* **18** (1990) 1158-1171.

- [16] A. Le Breton, On continuous and discrete sampling for parameter estimation in diffusion type processes. *Math. Programming Studies* **5** (1976) 124-144.
- [17] R.S. Liptser, A.N. Shiryaev, *Statistics of Random Processes: II Applications*, Second Edition, *Applications of Mathematics*, Springer-Verlag, Berlin, Heidelberg, New York, 2001.
- [18] A.W. Lo, Maximum likelihood estimation of generalized Ito processes with discretely sampled data, *Econometric Theory* **4** (1988) 231-247.
- [19] H. Long, Least squares estimator for discretely observed Ornstein-Uhlenbeck processes with small Lévy noises, *Statistics and Probability Letters* **79** (2009) 2076-2085.
- [20] H. Long, Parameter estimation for a class of stochastic differential equations driven by small stable noises from discrete observations, *Acta Mathematica Scientia* **30B** (2010) 645-663.
- [21] C. Ma, A note on “Least squares estimator for discretely observed Ornstein-Uhlenbeck processes with small Lévy noises”, *Statistics and Probability Letters* **80** (2010) 1528-1531.
- [22] H. Masuda, Simple estimators for parametric Markovian trend of ergodic processes based on sampled data, *J. Japan Statist. Soc.* **35** (2005) 147-170.
- [23] H. Masuda, Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein-Uhlenbeck processes, *Electronic Journal of Statistics*, **4** (2010) 525-565.
- [24] T. Ogihara, N. Yoshida, Quasi-likelihood analysis for the stochastic differential equation with jumps, *Stat. Inference Stoch. Process.* **14** (2011) 189-229.
- [25] A.R. Pedersen, A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations, *Scand. J. Statist.* **22** (1995) 55-71.
- [26] A.R. Pedersen, Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes, *Bernoulli* **1** (1995) 257-279.
- [27] R. Poulsen, Approximate maximum likelihood estimation of discretely observed diffusion processes, *Tech. Report 29*, Centre for Analytical Finance, University of Aarhus, 1999.
- [28] B.L.S. Prakasa Rao, Asymptotic theory for nonlinear least squares estimator for diffusion processes, *Math. Operations forschung Statist Ser. Statist.* **14** (1983) 195-209.
- [29] B.L.S. Prakasa Rao, *Statistical Inference for Diffusion Type Processes*, Arnold, London, Oxford University Press, New York, 1999.

- [30] Y. Shimizu, M-estimation for discretely observed ergodic diffusion processes with infinite jumps, *Stat. Inference Stoch. Process.* **9** (2006) 179-225.
- [31] Y. Shimizu, Quadratic type contrast functions for discretely observed non-ergodic diffusion processes. Research Report Series 09-04, Division of Mathematical Science, Osaka University, 2010.
- [32] Y. Shimizu, N. Yoshida, Estimation of parameters for diffusion processes with jumps from discrete observations, *Stat. Inference Stoch. Process.* **9** (2006) 227-277.
- [33] M. Sørensen, Small dispersion asymptotics for diffusion martingale estimating functions, Preprint No. 2000-2, Department of Statistics and Operation Research, University of Copenhagen, Copenhagen, 2000.
- [34] H. Sørensen, Parameter inference for diffusion processes observed at discrete points in time: a survey, *Internat. Statist. Rev.* **72** (2004) 337-354.
- [35] M. Sørensen, M. Uchida, Small diffusion asymptotics for discretely sampled stochastic differential equations, *Bernoulli* **9** (2003) 1051-1069.
- [36] K. Spiliopoulos, Methods of moments estimation of Ornstein-Uhlenbeck processes driven by general Lévy process, Preprint, University of Maryland, 2008.
- [37] A. Takahashi, An asymptotic expansion approach to pricing contingent claims, *Asia-Pacific Financial Markets* **6** (1999) 115-151.
- [38] A. Takahashi, N. Yoshida, An asymptotic expansion scheme for optimal investment problems, *Stat. Inference Stoch. Process.* **7** (2004) 153-188.
- [39] M. Uchida, Estimation for discretely observed small diffusions based on approximate martingale estimating functions, *Scand. J. Statist.* **31** (2004) 553-566.
- [40] M. Uchida, Approximate martingale estimating functions for stochastic differential equations with small noises, *Stochastic Process. Appl.* **118** (2008) 1706-1721.
- [41] M. Uchida, N. Yoshida, Information criteria for small diffusions via the theory of Malliavin-Watanabe, *Stat. Inference Stoch. Process.* **7** (2004) 35-67.
- [42] M. Uchida, N. Yoshida, Asymptotic expansion for small diffusions applied to option pricing, *Stat. Inference Stoch. Process.* **7** (2004) 189-223.
- [43] L. Valdivieso, W. Schoutens, F. Tuerlinckx, Maximum likelihood estimation in processes of Ornstein-Uhlenbeck type, *Stat. Infer. Stoch. Process.* **12** (2009) 1-19.
- [44] A.W. van der Vaart, *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics 3, Cambridge University Press, 1998.

- [45] N. Yoshida, Asymptotic expansion of maximum likelihood estimators for small diffusions via the theory of Malliavin-Watanabe, *Probab. Theory Relat. Fields* **92** (1992) 275-311.
- [46] N. Yoshida, Asymptotic expansion for statistics related to small diffusions, *J. Japan Statist. Soc.* **22** (1992) 139-159.
- [47] N. Yoshida, Conditional expansions and their applications, *Stochastic Process. Appl.* **107** (2003) 53-81.